

AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

- 1 1. (Currently amended) A method for characterizing a document with
2 respect to clusters of conceptually related words, comprising:
3 receiving the document, wherein the document contains a set of words;
4 selecting candidate clusters of conceptually related words that are related
5 to the set of words;
6 wherein the candidate clusters are selected using a model that explains
7 how sets of words are generated from clusters of conceptually related words,
8 wherein the conceptually related words are words that relate to ~~a single idea~~ a
9 common topic; and
10 constructing a set of components to characterize the document, wherein
11 the set of components includes components for candidate clusters, wherein each
12 component indicates a degree to which a corresponding candidate cluster is
13 related to the set of words,
14 ~~wherein the set of components is subsequently used to generate a response~~
15 ~~to a query from a user.~~
16 wherein the set of components provides an abstract representation for the
17 document, wherein the abstract representation is subsequently used as a substitute
18 for the document during query operations involving the document.

1 2. (Original) The method of claim 1, wherein the model is a probabilistic
2 model, which contains nodes representing random variables for words and for
3 clusters of conceptually related words.

1 3. (Original) The method of claim 2, wherein each component in the set of
2 components indicates a degree to which a corresponding candidate cluster is
3 active in generating the set of words.

1 4. (Original) The method of claim 3,
2 wherein nodes in the probabilistic model are coupled together by weighted
3 links; and
4 wherein if a cluster node in the probabilistic model fires, a weighted link
5 from the cluster node to another node can cause the other node to fire.

1 5. (Original) The method of claim 4, wherein if a node has multiple parent
2 nodes that are active, the probability that the node does not fire is the product of
3 the probabilities that links from the active parent nodes do not fire.

1 6. (Original) The method of claim 2, wherein the probabilistic model
2 includes a universal node that is always active and that has weighted links to all
3 cluster nodes.

1 7. (Original) The method of claim 4, wherein selecting the candidate
2 clusters involves:
3 constructing an evidence tree by starting with terminal nodes associated
4 with the set of words in the document, and following links in the reverse direction
5 to parent cluster nodes;

6 using the evidence tree to estimate a likelihood that each parent cluster
7 node was active in generating the set of words; and
8 selecting a parent cluster node to be a candidate cluster node based on its
9 estimated likelihood.

1 8. (Original) The method of claim 7, wherein estimating the likelihood that
2 a given parent node is active in generating the set of words may involve
3 considering:

4 the unconditional probability that the given parent node is active;
5 conditional probabilities that the given parent node is active assuming
6 parent nodes of the given parent node are active; and
7 conditional probabilities that the given parent node is active assuming
8 child nodes of the given parent node are active.

1 9. (Original) The method of claim 8, wherein considering the conditional
2 probabilities involves considering weights on links between nodes.

1 10. (Original) The method of claim 7 wherein estimating the likelihood
2 that a given parent node is active in generating the set of words involves marking
3 terminal nodes during the estimation process to ensure that terminal nodes are not
4 factored into the estimation more than once.

1 11. (Original) The method of claim 7, wherein constructing the evidence
2 tree involves pruning unlikely nodes from the evidence tree.

1 12. (Original) The method of claim 3, wherein during construction of the
2 set of components, the degree to which a candidate cluster is active in generating

3 the set of words is determined by calculating a probability that a candidate cluster
4 is active in generating the set of words.

1 13. (Original) The method of claim 3, wherein during construction of the
2 set of components, the degree to which a candidate cluster is active in generating
3 the set of words is determined by multiplying a probability that a candidate cluster
4 is active in generating the set of words by an activation for the candidate cluster,
5 wherein the activation indicates how many links from the candidate cluster to
6 other nodes are likely to fire.

1 14. (Original) The method of claim 1, wherein constructing the set of
2 components involves normalizing the set of components.

1 15. (Original) The method of claim 3, wherein constructing the set of
2 components involves approximating a probability that a given candidate cluster is
3 active over states of the probabilistic model that could have generated the set of
4 words.

1 16. (Original) The method of claim 15, wherein approximating the
2 probability involves:
3 selecting states for the probabilistic model that are likely to have generated
4 the set of words in the document; and
5 considering only selected states while calculating the probability that the
6 given candidate cluster is active.

1 17. (Original) The method of claim 16, wherein selecting a state that is
2 likely to have generated the set of words involves:
3 randomly selecting a starting state for the probabilistic model; and

4 performing hill-climbing operations beginning at the starting state to reach
5 a state that is likely to have generated the set of words.

1 18. (Original) The method of claim 17, wherein performing the hill-
2 climbing operations involves periodically changing states of individual candidate
3 clusters without regards to an objective function for the hill-climbing operations
4 to explore states of the probabilistic model that are otherwise unreachable through
5 hill-climbing operations.

1 19. (Original) The method of claim 18, wherein changing a state of an
2 individual candidate cluster involves temporarily fixing the changed state to
3 produce a local optimum for the objective function, which includes the changed
4 state.

1 20. (Original) The method of claim 1, wherein the document can include:
2 a web page; or
3 a set of terms from a query.

1 21. (Currently amended) A computer-readable storage medium storing
2 instructions that when executed by a computer cause the computer to perform a
3 method for characterizing a document with respect to clusters of conceptually
4 related words, wherein the computer-readable storage medium is one of a disk
5 drive, a magnetic tape, a CDs (compact discs), and a DVDs (digital versatile disc
6 or digital video disc), the method comprising:
7 receiving the document, wherein the document contains a set of words;
8 selecting candidate clusters of conceptually related words that are related
9 to the set of words, wherein the conceptually related words are words that relate to
10 ~~a single idea~~ a common topic;

11 wherein the candidate clusters are selected using a model that explains
12 how sets of words are generated from clusters of conceptually related words; and
13 constructing a set of components to characterize the document, wherein
14 the set of components includes components for candidate clusters, wherein each
15 component indicates a degree to which a corresponding candidate cluster is
16 related to the set of words,

17 ~~wherein the set of components is subsequently used to generate a response~~
18 ~~to a query from a user.~~

19 wherein the set of components provides an abstract representation for the
20 document, wherein the abstract representation is subsequently used as a substitute
21 for the document during query operations involving the document.

1 22. (Original) The computer-readable storage medium of claim 21,
2 wherein the model is a probabilistic model, which contains nodes representing
3 random variables for words and for clusters of conceptually related words.

1 23. (Original) The computer-readable storage medium of claim 22,
2 wherein each component in the set of components indicates a degree to which a
3 corresponding candidate cluster is active in generating the set of words.

1 24. (Original) The computer-readable storage medium of claim 23,
2 wherein nodes in the probabilistic model are coupled together by weighted
3 links; and
4 wherein if a cluster node in the probabilistic model fires, a weighted link
5 from the cluster node to another node can cause the other node to fire.

1 25. (Original) The computer-readable storage medium of claim 24,
2 wherein if a node has multiple parent nodes that are active, the probability that the

3 node does not fire is the product of the probabilities that links from the active
4 parent nodes do not fire.

1 26. (Original) The computer-readable storage medium of claim 22,
2 wherein the probabilistic model includes a universal node that is always active
3 and that has weighted links to all cluster nodes.

1 27. (Original) The computer-readable storage medium of claim 24,
2 wherein selecting the candidate clusters involves:
3 constructing an evidence tree by starting with terminal nodes associated
4 with the set of words in the document, and following links in the reverse direction
5 to parent cluster nodes;
6 using the evidence tree to estimate a likelihood that each parent cluster
7 node was active in generating the set of words; and
8 selecting a parent cluster node to be a candidate cluster node based on its
9 estimated likelihood.

1 28. (Original) The computer-readable storage medium of claim 27,
2 wherein estimating the likelihood that a given parent node is active in generating
3 the set of words may involve considering:
4 the unconditional probability that the given parent node is active;
5 conditional probabilities that the given parent node is active assuming
6 parent nodes of the given parent node are active; and
7 conditional probabilities that the given parent node is active assuming
8 child nodes of the given parent node are active.

1 29. (Original) The computer-readable storage medium of claim 28,
2 wherein considering the conditional probabilities involves considering weights on
3 links between nodes.

1 30. (Original) The computer-readable storage medium of claim 27,
2 wherein estimating the likelihood that a given parent node is active involves
3 marking terminal nodes during the estimation process to ensure that terminal
4 nodes are not factored into the estimation more than once.

1 31. (Original) The computer-readable storage medium of claim 27,
2 wherein constructing the evidence tree involves pruning unlikely nodes from the
3 evidence tree.

1 32. (Original) The computer-readable storage medium of claim 23,
2 wherein during construction of the set of components, the degree to which a
3 candidate cluster is active in generating the set of words is determined by
4 calculating a probability that a candidate cluster is active in generating the set of
5 words.

1 33. (Original) The computer-readable storage medium of claim 23,
2 wherein during construction of the set of components, the degree to which a
3 candidate cluster is active in generating the set of words is determined by
4 multiplying a probability that a candidate cluster is active in generating the set of
5 words by an activation for the candidate cluster, wherein the activation indicates
6 how many links from the candidate cluster to other nodes are likely to fire.

1 34. (Original) The computer-readable storage medium of claim 21,
2 wherein constructing the set of components involves normalizing the set of
3 components.

1 35. (Original) The computer-readable storage medium of claim 23,
2 wherein constructing the set of components involves approximating a probability
3 that a given candidate cluster is active over states of the probabilistic model that
4 could have generated the set of words.

1 36. (Original) The computer-readable storage medium of claim 35,
2 wherein approximating the probability involves:
3 selecting states for the probabilistic model that are likely to have generated
4 the set of words in the document; and
5 considering only selected states while calculating the probability that the
6 given candidate cluster is active.

1 37. (Original) The computer-readable storage medium of claim 36,
2 wherein selecting a state that is likely to have generated the set of words involves:
3 randomly selecting a starting state for the probabilistic model; and
4 performing hill-climbing operations beginning at the starting state to reach
5 a state that is likely to have generated the set of words.

1 38. (Original) The computer-readable storage medium of claim 37,
2 wherein performing the hill-climbing operations involves periodically changing
3 states of individual candidate clusters without regards to an objective function for
4 the hill-climbing operations to explore states of the probabilistic model that are
5 otherwise unreachable through hill-climbing operations.

1 39. (Original) The computer-readable storage medium of claim 38,
2 wherein changing a state of an individual candidate cluster involves temporarily
3 fixing the changed state to produce a local optimum for the objective function,
4 which includes the changed state.

1 40. (Original) The computer-readable storage medium of claim 21,
2 wherein the document can include:
3 a web page; or
4 a set of terms from a query.

1 41. (Currently amended) An apparatus for characterizing a document with
2 respect to clusters of conceptually related words, comprising:
3 a receiving mechanism, configured to receive the document, wherein the
4 document contains a set of words;
5 a selection mechanism configured to select candidate clusters of
6 conceptually related words that are related to the set of words;
7 wherein the candidate clusters are selected using a model that explains
8 how sets of words are generated from clusters of conceptually related words,
9 wherein the conceptually related words are words that relate to ~~a single idea~~ a
10 common topic; and
11 a component construction mechanism configured to construct a set of
12 components to characterize the document, wherein the set of components includes
13 components for candidate clusters, wherein each component indicates a degree to
14 which a corresponding candidate cluster is related to the set of words,
15 ~~wherein the set of components is subsequently used by a generation~~
16 ~~mechanism to generate a response to a query from a user.~~

17 wherein the set of components provides an abstract representation for the
18 document, wherein the abstract representation is subsequently used as a substitute
19 for the document during query operations involving the document.

1 42. (Original) The apparatus of claim 41, wherein the model is a
2 probabilistic model, which contains nodes representing random variables for
3 words and for clusters of conceptually related words.

1 43. (Original) The apparatus of claim 42, wherein each component in the
2 set of components indicates a degree to which a corresponding candidate cluster is
3 active in generating the set of words.

1 44. (Original) The apparatus of claim 43,
2 wherein nodes in the probabilistic model are coupled together by weighted
3 links; and
4 wherein if a cluster node in the probabilistic model fires, a weighted link
5 from the cluster node to another node can cause the other node to fire.

1 45. (Original) The apparatus of claim 44, wherein if a node has multiple
2 parent nodes that are active, the probability that the node does not fire is the
3 product of the probabilities that links from the active parent nodes do not fire.

1 46. (Original) The apparatus of claim 43, wherein the probabilistic model
2 includes a universal node that is always active and that has weighted links to all
3 cluster nodes.

1 47. (Original) The apparatus of claim 44, wherein the selection mechanism
2 is configured to:

3 construct an evidence tree by starting with terminal nodes associated with
4 the set of words in the document, and following links in the reverse direction to
5 parent cluster nodes;
6 use the evidence tree to estimate a likelihood that each parent cluster node
7 was active in generating the set of words; and to
8 select a parent cluster node to be a candidate cluster node based on its
9 estimated likelihood.

1 48. (Original) The apparatus of claim 47, wherein while estimating the
2 likelihood that a given parent node is active in generating the set of words, the
3 selection mechanism is configured to consider at least one of the following:
4 the unconditional probability that the given parent node is active;
5 conditional probabilities that the given parent node is active assuming
6 parent nodes of the given parent node are active; and
7 conditional probabilities that the given parent node is active assuming
8 child nodes of the given parent node are active.

1 49. (Original) The apparatus of claim 48, wherein while considering the
2 conditional probabilities, the selection mechanism is configured to consider
3 weights on links between nodes.

1 50. (Original) The apparatus of claim 47, wherein while estimating the
2 likelihood that a given parent node is active in generating the set of words, the
3 selection mechanism is configured to mark terminal nodes during the estimation
4 process to ensure that terminal nodes are not factored into the estimation more
5 than once.

1 51. (Original) The apparatus of claim 47, wherein while constructing the
2 evidence tree, the selection mechanism is configured to prune unlikely nodes from
3 the evidence tree.

1 52. (Original) The apparatus of claim 43, wherein while constructing a
2 given component in the set of components, the component construction
3 mechanism is configured to determine the degree to which a candidate cluster is
4 active in generating the set of words by calculating a probability that a candidate
5 cluster is active in generating the set of words.

1 53. (Original) The apparatus of claim 43, wherein while constructing a
2 given component in the set of components, the component construction
3 mechanism is configured to determine the degree to which a candidate cluster is
4 active in generating the set of words by multiplying a probability that a candidate
5 cluster is active in generating the set of words by an activation for the candidate
6 cluster, wherein the activation indicates how many links from the candidate
7 cluster to other nodes are likely to fire.

1 54. (Original) The apparatus of claim 41, wherein the component
2 construction mechanism is configured to normalize the set of components.

1 55. (Original) The apparatus of claim 43, wherein the component
2 construction mechanism is configured to approximate a probability that a given
3 candidate cluster is active over states of the probabilistic model that could have
4 generated the set of words.

1 56. (Original) The apparatus of claim 55, wherein while approximating the
2 probability, the component construction mechanism is configured to:

3 select states for the probabilistic model that are likely to have generated
4 the set of words in the document; and to
5 consider only selected states while calculating the probability that the
6 given candidate cluster is active.

1 57. (Original) The apparatus of claim 56, wherein while selecting a state
2 that is likely to have generated the set of words, the component construction
3 mechanism is configured to:

4 randomly select a starting state for the probabilistic model; and to
5 perform hill-climbing operations beginning at the starting state to reach a
6 state that is likely to have generated the set of words.

1 58. (Currently amended) The apparatus of claim 57, wherein while
2 performing the hill-climbing operations, the component construction mechanism
3 is configured to periodically change states of individual candidate clusters without
4 regards to an objective function for the hill-climbing operations to explore states
5 of the probabilistic model that are otherwise unreachable through hill-climbing
6 operations.

1 59. (Original) The apparatus of claim 58, wherein while changing a state
2 of an individual candidate cluster, the component construction mechanism is
3 configured to temporarily fix the changed state to produce a local optimum for the
4 objective function, which includes the changed state.

1 60. (Original) The apparatus of claim 41, wherein the document can
2 include:
3 a web page; or
4 a set of terms from a query.

1 61. (Currently amended) A computer-readable storage medium containing
2 a data structure that facilitates characterizing a document with respect to clusters
3 of conceptually related words, ~~wherein the computer-readable storage medium is~~
4 ~~one of a disk drive, a magnetic tape, a CDs (compact discs), and a DVDs (digital~~
5 ~~versatile disc or digital video disc)~~, the data structure comprising:
6 a probabilistic model that contains nodes representing random variables
7 for words and for clusters of conceptually related words, wherein the conceptually
8 related words are words that relate to ~~a single idea~~ a common topic;
9 wherein nodes in the probabilistic model are coupled together by weighted
10 links;
11 wherein if a cluster node in the probabilistic model fires, a weighted link
12 from the cluster node to another node can cause the other node to fire; and
13 wherein the other code can be associated with a word or a cluster.

1 62. (Original) The computer-readable storage medium of claim 61,
2 wherein the probabilistic model includes a universal node that is always active
3 and that has weighted links to all cluster nodes.